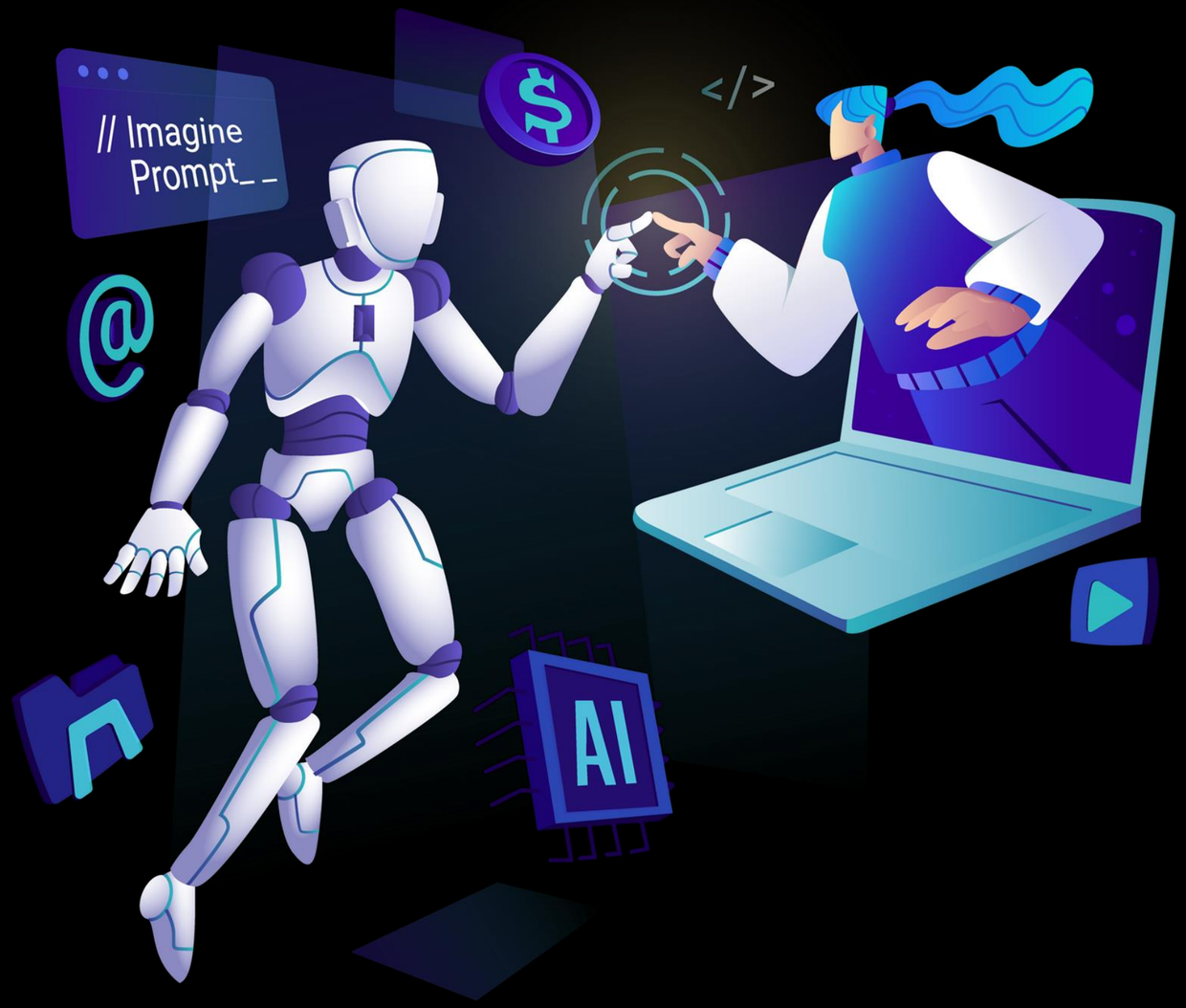


UNSUPERVISED LEARNING

CLUSTERING & PRINCIPAL COMPONENT ANALYSIS (PCA)

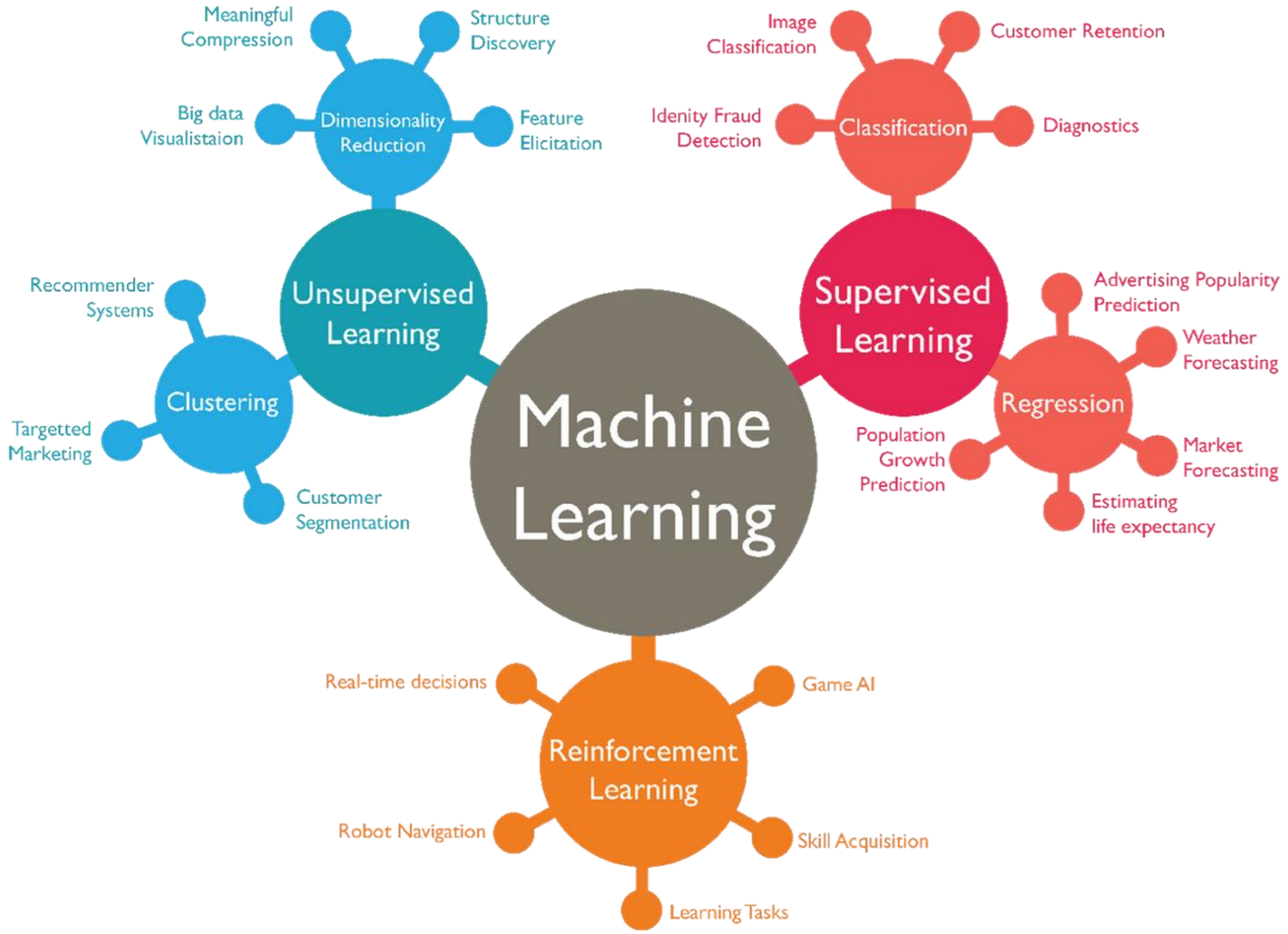
ANJANI KASHYAP (109) & NAYAN KASTURI (110)

Content Outline

Topics for discussion

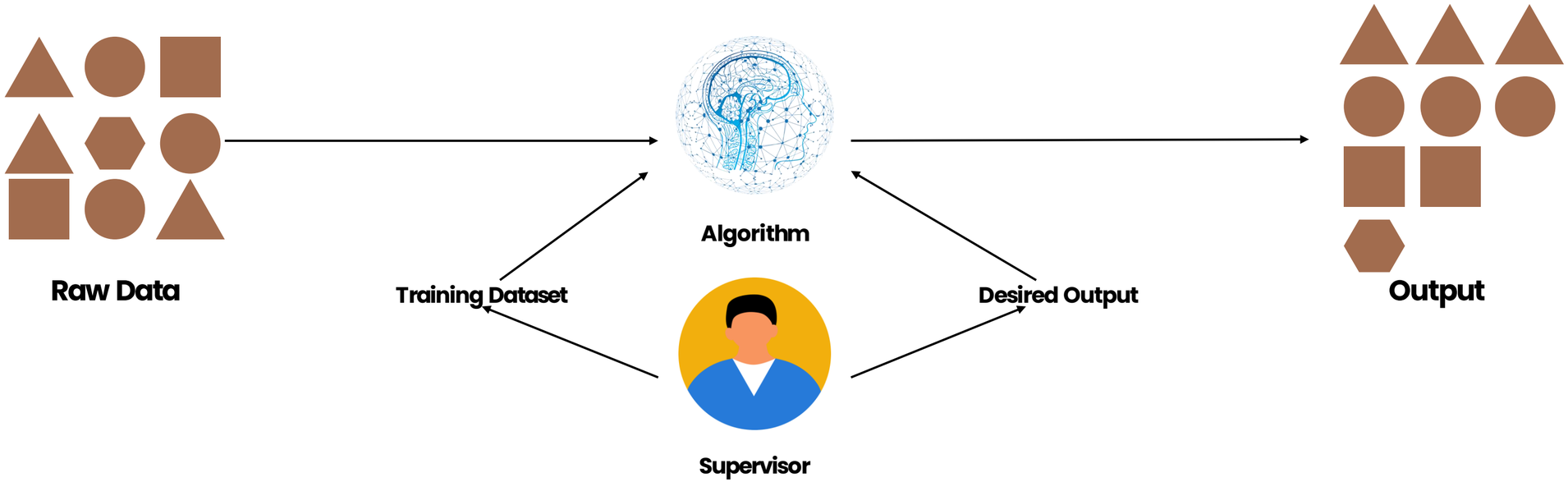
01. Machine Learning

- **What is Machine Learning or ML?**
- **How does ML works?**
- **What are the applications of ML?**

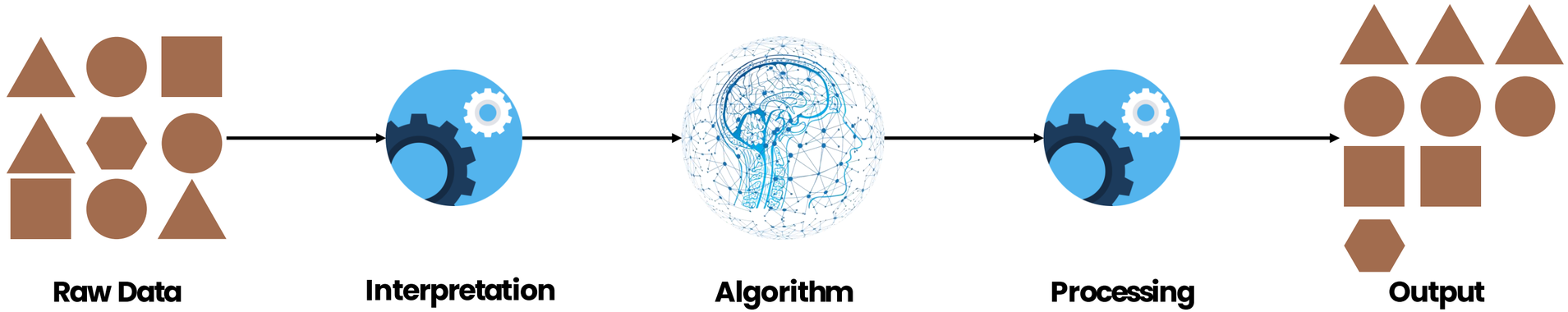


02. Supervised v/s Unsupervised Learning

SUPERVISED LEARNING



UNSUPERVISED LEARNING



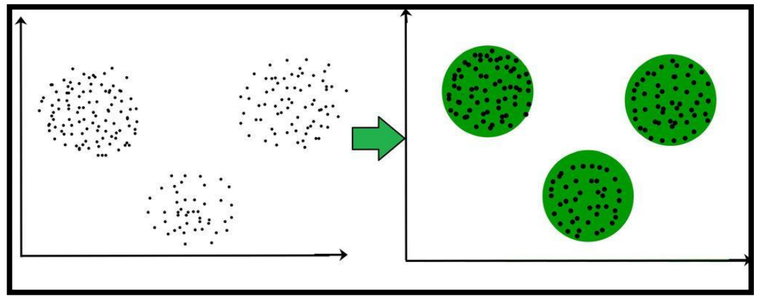
03. Unsupervised Learning

- **What is Unsupervised Learning in ML?**
- **Process and Working of Unsupervised Learning Models.**
- **Applications of Unsupervised Learning**

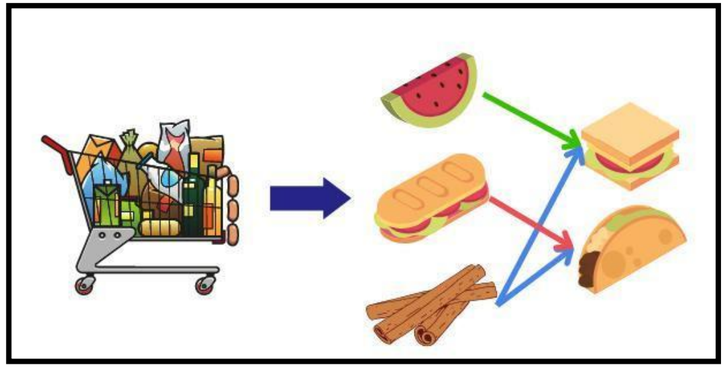
Types of Unsupervised Learning:

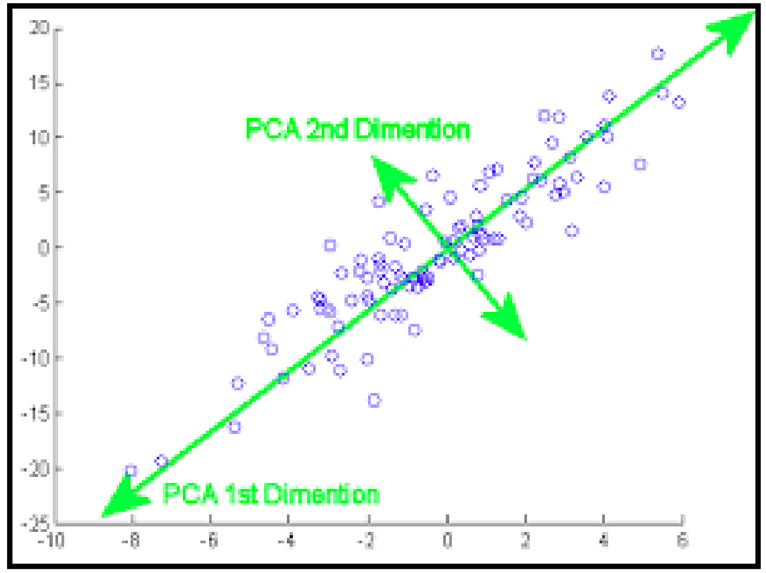
Dimensionality Reduction & Principal Component Analysis (PCA)

Clustering



Association





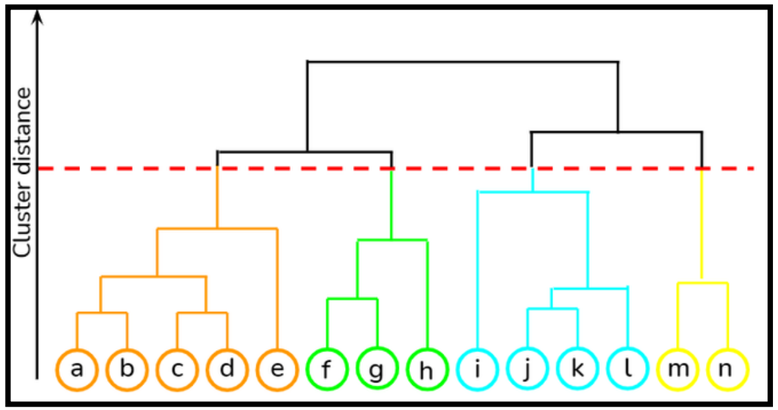
04. Clustering

- **What is a clustering? What are its importance?**

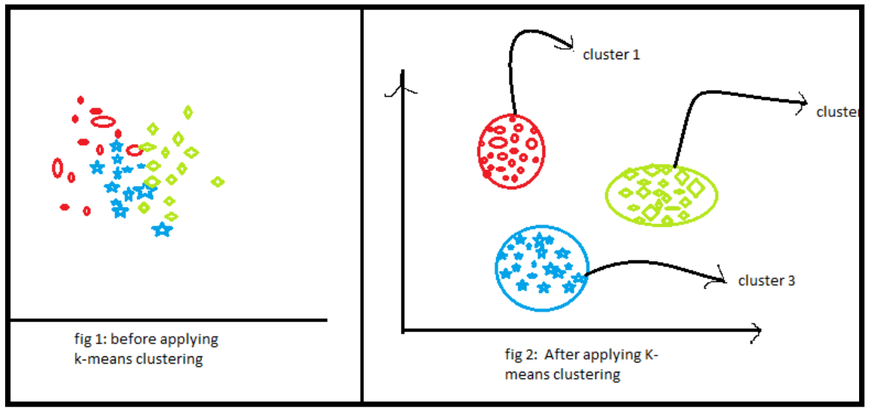
- **Some common clustering algorithms/types:**
  - K-means clustering
  - Hierarchical Clustering
  - DBSCAN Clustering
  - Normal Distribution
  - Spectral Clustering

05. Types of Clustering Methods

Hierarchical Clustering: It builds a hierarchy of clusters by successively merging or splitting them. This approach creates a tree-like structure (dendrogram) that shows how data can be grouped at different levels.



K-Means Clustering: This is one of the most popular clustering algorithms. It groups data into 'K' clusters, where K is a user-defined parameter. The algorithm tries to minimize the distance between data points in the same cluster and maximize the distance between different clusters.



06. Applications of Clustering

Marketing: To identify groups of households that are similar to each other based on factors such as household income, household size, and occupation of the head of the household.
Streaming services: To analyze viewer's behaviour and suggest them videos and show them relevant advertisements.
Sports science: To predict the match results and other analysis.
Document analysis: To organize documents quickly and efficiently and also to create datasets using the content in the documents.
Finance: To determine whether a banknote is authentic or fake. A K-Means Clustering model can be developed to classify banknotes based on their features.

Clustering has many other applications in field of bioinformatics as well, such as for, image processing, and pattern recognition. It is a powerful tool for data analysis and can help in identifying patterns and relationships in data that may not be immediately apparent.

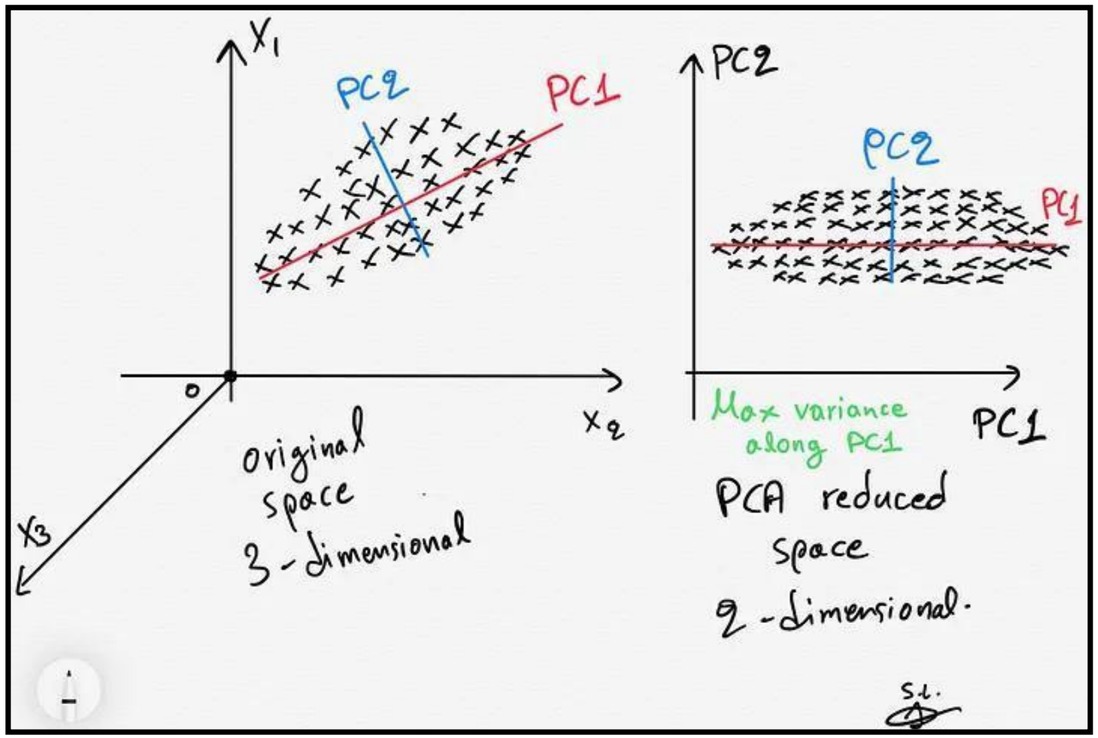
07. Principal Component Analysis (PCA)

- **What is PCA? Does it reduces dimensionality randomly?**

- **How does it actually works?**

Important features of PCA:

- Its is used to reduce the dimensionality of the dataset.
- It is used to identify and analyze the important features in a particular dataset.
- It also helps in determining the number of principal components that needs to be retained in the reduced dataset



08. Applications of PCA

Finance: To reduce the dimensionality of financial data and identify the important factors that drive asset prices, thus helping in portfolio optimization, risk management, and asset pricing.

Image processing: PCA is used for image compression as well as for image recognition and object detection.

Neuroscience: To analyze electroencephalography (EEG) data and ultimately to measure the activity of neurons in the brain and identify patterns in the data.

Medical imaging: PCA's feature extraction and dimensionality reduction is used in identifying important features in medical images and reducing size of dataset.

PCA has many other applications in various fields such as marketing, healthcare, and sports science. It is a powerful tool for data analysis and can help in identifying patterns and relationships in data that may not be immediately apparent.

09. Conclusion

Unsupervised learning with PCA and clustering helps simplify complex data, discover patterns, and gain insights. In bioinformatics and data analysis, these techniques are essential for understanding unstructured data and making better decisions. Embracing unsupervised learning is the key to unlocking valuable information from unlabeled data and improving our understanding of complex systems.

10. Q&A and References

References:

- Tucci, L., & Burns, E. (n.d.). *What is machine learning and how does it work? In-depth guide.* https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML
- Jaadi, Z., Whitfield, B., & Pierre, S. (2023, March 29). *A Step-by-Step Explanation of Principal Component Analysis (PCA).* https://builtin.com/data-science/step-step-explanation-principal-component-analysis
- Alam, M. (2020, December 11). *Clustering: Concepts, algorithms and applications.* https://towardsdatascience.com/clustering-concepts-algorithms-and-applications-f512a949549a